BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2024
http://excel.fit.vutbr.cz

Deep Learning for 3D Mesh Registration

Student: Bc. Dávid Pukanec
Supervisor: Ing. Michal Španěl, Ph.D.

Abstract

The problem of mesh alignment is often solved through point cloud registration. The registration task faces several challenges e.g., noise, partial data, and sparse density. Numerous deep learning-based registration methods are published every year achieving state-of-the-art results. Based on their core concepts, the methods can loosely be divided into correspondence-based and correspondence-free. Even though comparisons of individual methods exist, the cross evaluations of both categories are lacking. In this work, we present a deeper evaluation of *Lepard* and *FINet*, on the ModelNet dataset. We show that *FINet* outperforms *Lepard* in low-density clouds by 6° Error(R) and 0.0077 Error(t). Moreover, we show that both methods are sensitive to the data preparation process.

*xpukan02@stud.fit.vutbr.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

Registering 3D models is a longstanding and crucial task in computer vision. The primary objective is to estimate the rotation and translation between *source* and *target* models. These models are often represented in the form of point clouds, as shown in Introduction section of the poster. Conventional methods such as ICP tend to fall into local optimum and struggle with more challenging tasks when the point clouds are obscured with noise and partiality. The most common registration uses are in augmented reality, medical imaging, 3D reconstruction, and autonomous driving.

Recently, many deep learning methods have been developed, which can be categorised into correspondence-based and correspondence-free. Despite this very little attention is given to the cross-evaluation of methods in those two categories. Moreover, these methods are often evaluated on different datasets using different metrics.

This work examines two methods called *Lepard* and *FINet*. We test both methods on a set of experiments using the same data and evaluation metrics. Furthermore, we created our own sampling and augmentation pipeline following the works of [1].

2. Evaluated Method

Lepard is a correspondence-based method. It is designed around the KPFCN feature extractor, the concept of Transformers with self-attention and cross-attention, and differentiable matching. As this method outputs matched features, it uses RANSAC to estimate the final transformation. We can see the architecture in Figure 1 section. The authors introduced three novel ideas to leverage this 3D positional information [2].

- First, a network that fully disentangles the point cloud representation into a feature space and a position space.
- Second, a positional encoding function is used, that reveals relative distance information between points.
- Thirdly, a repositioning module that applies transformation in between transformer layers to enhance the relative position between the source and *target* point *cloud*.

FINet falls into the correspondence-free category. It directly regresses the rotation and translation from the extracted features. It emphasises data association as a significant step in registering models. This approach introduces three innovative ideas listed below [3]:

- First, an embedding module is proposed that manages to extract multi-level features. This promotes the data association between inputs.
- Secondly, the feature extraction is split into two separate branches, as shown in Figure 2. One for translation and the second for rotation feature extraction.
- Finally, the transformation-sensitive loss is introduced, which enhances the extraction of rotation-attentive and translation-attentive features.

3. Data

ModelNet40 dataset is used to evaluate both methods. It consists of 40 categories with 12 311 CAD models. To prepare the data for training both networks, we constructed a pipeline similar to the works of [4]. We sample point clouds from given meshes using uniform sampling.

To introduce challenging scenarios, as illustrated in Challenges, we created a data augmentation pipeline inspired by works of [1]. We augment the data just before running it through the network. Some of the augmentations are Gaussian noise, random rotation and translation, and partiality.

4. Metrics

The basic metrics can be split into anisotropic and isotropic. For anisotropic we follow the work of [5]. These metrics are Mean Absolute Error and Mean Square Error for rotation and translation. Due to space, we only display the isotropic errors, such as Mean Isotropic Rotation and Translation Errors, because they provide a clearer image of the misalignment, see equation 1.

$$Error(\mathbf{R}) = \angle(\mathbf{R}_{GT}^{-1}\hat{\mathbf{R}}) \text{ ; } Error(\mathbf{t}) = \|t_{GT} - \hat{t}\|_2 \quad (1)$$

Where $\{\mathbf{R}_{GT}, t_{GT}\}$ is ground truth and $\{\hat{\mathbf{R}}, \hat{t}\}$ is predicted rotation and tranlsation. The $\angle$ returns the rotation angle of the matrix in degrees [1]. Meanwhile, the translation error is displayed in proportion to the unit sphere.

Above mentioned metrics unfairly penalize for symmetries, present in the samples. Therefore, we also calculate the Chamfer distance metric between the *source* point cloud and the *target* point cloud.

5. Experiments

The first experiment showed in Results section of the poster is focused on highlighting the correspondence impact in the dataset. With our pipeline, we constructed three distinct datasets:

- **Once sampled** data **without** subsampling (**OS/w**). Exact one-to-one correspondence.
- **Once sampled** data **with** subsampling (**OS**). Varying amount of one-to-one correspondence.
- **Twice sampled** data **with** subsampling (**TS**). Zero one to one-to-one correspondence.

The results can be seen in the first table of Table 1. Interestingly both methods are impacted with decreasing amounts of one-to-one correspondences, as we would suspect this only to happen with Lepard. A second observation is that FINets error increases significantly on unseen categories (test subset).

The second experiment Table 2 is dedicated to challenging cases of registration. Data was augmented with Gaussian noise and partiality of varying magnitude. The $\approx 72\%$ overlap is the usual testing norm. We further pushed the limits to $\approx 53\%$ overlap. It is worth noting that the worse performance of Lepard might be due to the sparse density of the point cloud.

Figure 3 is dedicated to the significant increase in the error of FINet, evaluated on unseen categories. The plot shows per category rotational error, where we can see that some categories e.g., plant, stairs, and stool, are much worse compared to Lepard. The person and radio category performs worse for both.

The visual representation of registered data is presented in Results section. We highlight several examples including the use of ICP or subsequent refinement with it.

6. Conclusions

This work showcases the abilities of both types of point cloud registration methods. We constructed our own data sampling and augmentation pipeline to perform given experiments. Under simple conditions, both methods perform well with Lepard performing slightly better. However, with added imperfection, the robustness of FINet is highlighted.

During the creation of this poster, we experimented with more dense point clouds. Moreover, we try to combat the FINet's increasing error in some categories. Unfortunately, due to time restrictions, we were not able to showcase our results.

References

[1] Zi Jian Yew and Gim Hee Lee. Rpm-net: Robust point matching using learned features. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11824–11833, 2020.

[2] Yang Li and Tatsuya Harada. Lepard: Learning partial point cloud matching in rigid and deformable scenes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5554–5564, 2022.

[3] Hao Xu, Nianjin Ye, Guanghui Liu, Bing Zeng, and Shuaicheng Liu. Finet: Dual branches feature interaction for partial-to-partial point cloud registration. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 2848–2856, 2022.

[4] Yasuhiro Aoki, Hunter Goforth, Rangaprasad Arun Srivatsan, and Simon Lucey. Pointnetlk: Robust & efficient point cloud registration using pointnet. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7163–7172, 2019.

[5] Yue Wang and Justin M Solomon. Deep closest point: Learning representations for point cloud registration. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3523–3532, 2019.